Sea Surface Temperature Forecasting Using Machine Learning

SAMTA KUMARI[1], SAJID ALI[2], SUSHANT RANJAN[3], DR. ISHRAT ALI[4], PROF. (DR.) SANJAY PACHAURI[5]
[1, 2, 3, 4, 5]*Department of Data Science (DDCS), GNIOT College, Greater Noida, India*

Abstract- Sea Surface Temperature (SST) is a critical climate variable that influences global weather, monsoon behavior, and marine ecosystems. Traditional numerical models struggle with high computational cost and nonlinear ocean–atmosphere dynamics. This work presents a machine-learning-based framework for SST forecasting using satellite observations and reanalysis data. Models including Random Forest, LSTM, and ConvLSTM are evaluated for short- and medium-term prediction. Results show that ML models significantly outperform persistence and statistical baselines in accuracy and efficiency. The study demonstrates the potential of data-driven methods to enhance operational SST forecasting and support climate monitoring applications.

Index Terms- Sea Surface Temperature (SST); Machine Learning; Deep Learning; ConvLSTM; LSTM; Climate Forecasting; Oceanography; SatelliteData; Time-Series Prediction.

I. INTRODUCTION

Sea Surface Temperature (SST) plays a major role in global climate regulation, monsoon systems, and the formation of extreme weather events. Accurate SST forecasting is essential for climate monitoring, marine ecosystem protection, and disaster preparedness. Traditional numerical and statistical models often struggle with nonlinear ocean–atmosphere interactions and require high computational resources. Machine learning (ML) offers a powerful alternative by learning complex temporal and spatial patterns directly from satellite and reanalysis datasets. Recent models such as LSTM and ConvLSTM have shown strong potential for improving SST prediction accuracy. This study investigates ML-based SST forecasting and evaluates multiple models to identify effective, scalable approaches for operational use.

II. METHODOLOGY

The proposed SST forecasting framework consists of four major stages: data preparation, feature engineering, model development, and performance evaluation. Historical SST data from NOAA ERSST, satellite observations (MODIS/AVHRR), and ERA5 reanalysis variables are collected and preprocessed through quality control, anomaly computation, normalization, and gap filling. Relevant features such as lagged SST values, wind speed, air temperature, and sea-level pressure are generated to support multivariate forecasting.

Multiple machine learning models are developed, including Random Forest, XGBoost, LSTM, GRU, and ConvLSTM architectures. Grid search and time-aware validation are used for hyperparameter optimization. Models are trained using a sliding-window approach and evaluated across several forecast horizons. Performance is assessed using RMSE, MAE, MAPE, and R^2 metrics. Finally, the best-performing models are compared against persistence and traditional statistical baselines to determine forecasting improvements.

III. RESULTS AND DISCUSSION

Experimental results show that machine learning models outperform traditional forecasting approaches for short- and medium-term SST prediction. LSTM and ConvLSTM models achieve the lowest RMSE and highest R^2 scores, demonstrating strong ability to capture temporal and spatio-temporal patterns. Tree-based models like Random Forest and XGBoost perform well for short lead times but decline in longer forecasts. The analysis also shows that ML models better detect SST anomalies and early warming signals. Overall, the findings confirm that data-driven methods provide more accurate and efficient SST forecasts compared to classical baselines.

IV. CONCLUSION

This study demonstrates that machine learning provides an effective and scalable solution for Sea Surface Temperature (SST) forecasting. Models such as LSTM and ConvLSTM outperform traditional statistical and persistence approaches by better capturing nonlinear patterns and spatio-temporal variability in ocean data. Tree-based models offer reliable short-term predictions, while deep learning architectures show superior performance for longer lead times. Overall, the results highlight the potential of ML-based methods to enhance operational SST prediction, support climate monitoring, and improve early warning systems for marine and weather-related events. Future work can integrate physics-based constraints and real-time observational data to further improve forecast accuracy and robustness.

REFERENCES

[1] NOAA National Centers for Environmental Information, "ERSST v5: Extended Reconstructed Sea Surface Temperature," https://www.ncei.noaa.gov/, accessed 2025.

[2] ECMWF, "ERA5 Reanalysis Dataset," https://www.ecmwf.int/, accessed 2025.

[3] NASA, "MODIS and AVHRR Sea Surface Temperature Products," https://oceancolor.gsfc.nasa.gov/.

[4] Rasp, M. Pritchard, and P. Gentine, "Deep learning to represent subgrid processes in climate models," *Proc. Natl. Acad. Sci.*, vol. 115, no. 39, pp. 9684–9689, 2018.

[5] P. Aggarwal et al., "Applications of Machine Learning in Climate Science," in *Climate Informatics*, Springer, 2021.